QUANTITATIVE STRUCTURE–RETENTION RELATIONSHIPS (QSRR) IN CHROMATOGRAPHY

R. Kaliszan, Medical University of Gdańsk, Gdańsk, Poland

Copyright © 2000 Academic Press

Introduction

To relate structure and chromatographic retention an approach is needed that lacks the rigour of thermodynamics but which provides otherwise inaccessible information. Such an approach is a combination of detailed models with certain thermodynamic concepts.

Linear free-energy relationships (LFER) may be regarded as linear relationships between the logarithms of the rate or equilibria constants for one reaction series and those for a second reaction series subjected to the same variation in reactant structure or reaction conditions. Retention parameters can be assumed to reflect the free-energy changes associated with the chromatographic distribution process. Accordingly, a chromatographic column can be treated as a 'free-energy transducer', translating differences in chemical potentials of analytes, arising from differences in their structure, into quantitative differences in retention parameters.

Assuming LFER it is possible to determine relative inputs of individual structural groups, fragments or features, to a property measured for a series of compounds in various chemical, physical, physicochemical and biological experiments. Such structural parameters (descriptors) can then be related to retention parameters.

The existence of LFER is normally proved statistically. The basic methodology of employing LFER to predict differences in pharmacological activity within a series of related agents was proposed in 1964 by Hansch and Fujita (QSAR, quantitative structure–activity relationships). Multiple regression analysis was applied in 1977 to chromatographic data (QSRR, quantitative structure–retention relationships). Other chemometric methods of data analysis have since been introduced to QSRR. QSRR are now one of the most extensively studied manifestations of LFER and, at the same time, the most common application of chemometrics.

Methodology and Goals of QSRR Analysis

To undertake QSRR studies two kinds of input data are needed (**Figure 1**). One is a set of quantitatively comparable retention data (dependent variable) for a sufficiently large (for statistical reason) set of analytes. The other is a set of quantities (independent variables) assumed to account for structural differences among the analytes being studied. Through the use of chemometric computational techniques, retention parameters are characterized in terms of various descriptors of analytes (and/or their combinations) or in terms of systematic knowledge extracted (learnt) from these descriptors.

To obtain statistically significant and physically meaningful QSRR, reliable input data are required and stringent mathematical analysis must be carried out. If this is not done, formally valid correlations may be developed for chemically invalid principles.

Once good QSRR have been obtained, one can exploit them for:

1. prediction of retention of a new analyte;
2. identification of the most informative structural descriptors possessing the highest retention prediction potency;
3. insight into the molecular mechanism of separation operating in individual chromatographic systems;
4. evaluation of physicochemical properties of analytes, e.g. their hydrophobicity (lipophilicity);
5. prediction of relative biological (pharmacological) activities within a set of drugs and other xenobiotics.

Retention Parameters for QSRR

The great advantage of the QSRR analysis over other quantitative structure–property relationship studies is that chromatography can readily produce a large amount of relatively precise and reproducible data. In a chromatographic process all conditions may be kept constant and hence the structure of an analyte becomes the single independent variable in the system.

The most commonly used retention parameter in gas chromatography is the Kováts retention index. When the adjusted retention times are used to calculate retention indices, parameters are obtained that

DropBooks

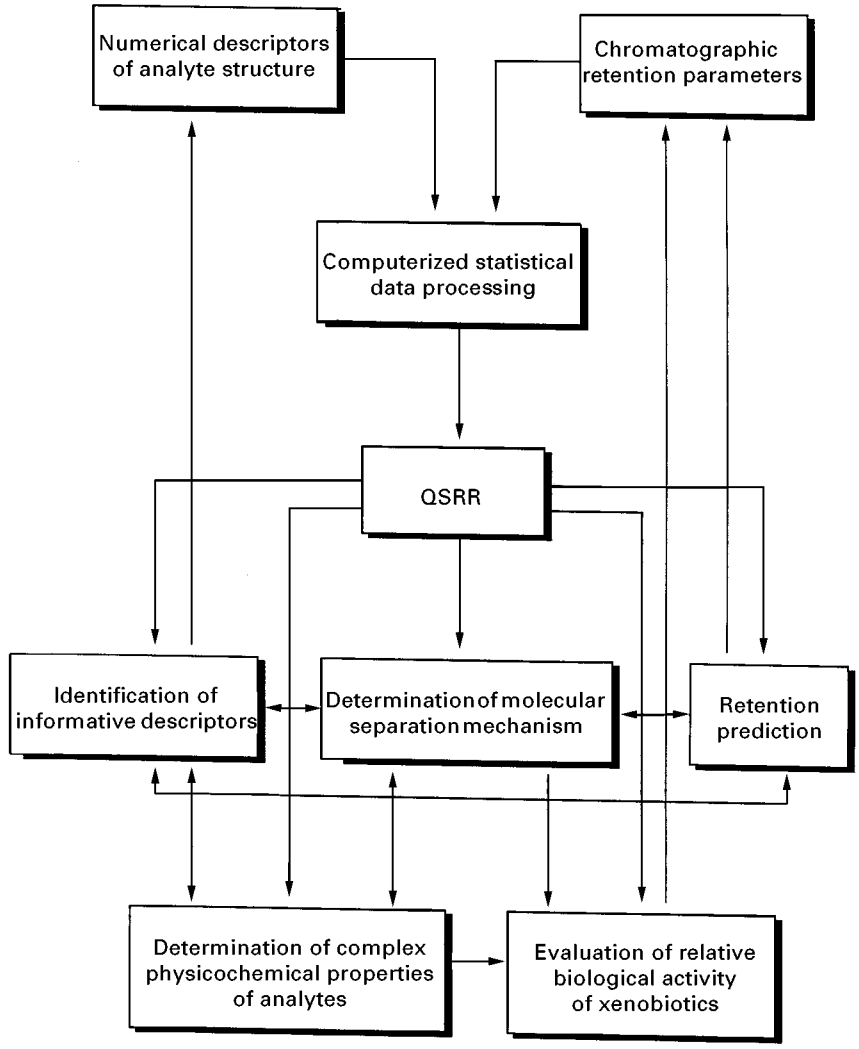

Numerical descriptors of analyte structure
Chromatographic retention parameters
Computerized statistical data processing
QSRR
Identification of informative descriptors
Determination of molecular separation mechanism
Retention prediction
Determination of complex physicochemical properties of analytes
Evaluation of relative biological activity of xenobiotics

Figure 1 Methodology and goals of studies of quantitative structure–retention relationships (QSRR). (Adapted with permission from Kaliszan R (1992) Quantitative structure–retention relationships. *Analytical Chemistry* 64: 619A–631A. Copyright 1992 American Chemical Society.)

depend only on the column temperature and the stationary phase used. Kováts retention indices are highly reproducible; with a well-designed experimental technique, an interlaboratory reproducibility of one unit is possible. Sometimes in QSRR studies the logarithms of retention volumes of solutes are used instead of Kováts indices.

Classical thin-layer chromatographic (TLC) retention parameters are of rather limited reproducibility. The use of well-defined small-diameter stationary phase particles and a better knowledge of the parameters that determine the efficiency of chromatographic systems have led to the development of high performance TLC (HPTLC). An advantage of TLC over column chromatography, from the point of view of QSRR studies, is that tens of analytes can be simultaneously chromatographed on the same plate. The retention parameter from TLC (and also from paper chromatography) that is normally used in QSRR is the R_M value. The R_M value is defined as log $((1/R_F) - 1)$, where R_F is the distance migrated by the sample from the origin compared with the distance migrated by the solvent front from the origin.

The LFER-based retention parameter in high performance liquid chromatography (HPLC) is the logarithm of the retention factor k. The retention factor is defined as in eqn [1].

$$k = (t_R - t_M)/t_M = (V_R - V_M)/V_M \qquad [1]$$

where t_R and V_R are the retention time and the retention volume of the analyte. The quantities t_M and V_M denote the elution time and the elution volume of an unretained compound.

HPLC retention data for QSRR analysis are usually obtained by measuring log k at several fixed eluent compositions (isocratic conditions) and then by extrapolating the dependence of log k on a binary eluent composition to a common mobile phase composition based on the Soczewiński–Snyder model:

$$\log k = \log k_w - S\varphi \quad [2]$$

In eqn [2] S is a constant for a given analyte and a given HPLC system and φ is the volume fraction of one of the mobile phase components. In the case of reversed-phase HPLC, k_w is a hypothetical retention factor expected for pure water (buffer) mobile phase ($\varphi = 0$).

The curvature often observed in plots of log k versus φ leads to a quadratic relationship:

$$\ln k = A\varphi^2 + B\varphi + C \quad [3]$$

where A, B and C are constants for a given analyte and a given chromatographic system. The ln k value calculated from eqn [3] by assuming $\varphi = 0$ is only occasionally used in QSRR analysis.

In spite of considerable effort, the relationships between retention and mobile phase composition are approximate. Often the values of log k_w extrapolated from a number of isocratic measurements in water/organic modifier eluents of varying compositions to a pure water eluent (the intercepts in eqn [2]) are different from those determined experimentally (when this is possible). Reversed-phase HPLC log k_w data are also usually different when derived from aqueous systems modified with different organic solvents. Still, the determination of log k_w appears to be the most reliable means of normalizing the retention parameters for QSRR analysis.

It should be noted here that some workers advocate using as the dependent variable the parameter S from eqn [2] or its ratio to log k_w.

The electrophoretic mobility, μ_{el}, of spherical particles is described by a simple equation:

$$\mu_{el} = (z\Phi)/(6\pi\eta aN) \quad [4]$$

where z is the effective charge, Φ is the charge per mole of protons, η is the viscosity of the medium, a is the radius of the charged species and N is the Avogadro number.

A parameter normally measured in capillary electrophoresis (CE) is migration time, t. In a given CE system this parameter is inversely proportional to the electrophoretic mobility, μ. A normalized parameter, μ ($cm^3\,V^{-1}$) allows comparison of data obtained in different CE systems. If a series of analytes is analysed under the same conditions, then $1/t$ and μ are equivalent.

Chemometric Procedures in QSRR

Assuming LFER, a given chromatographic retention parameter may be described (statistically) by a set of analyte structural descriptors:

$$\text{Retention parameter} = f(a_1x_1, \ldots, a_nx_n) \quad [5]$$

The coefficients a_1–a_n for individual n descriptors are calculated by multiple regression. There are computer programs available commercially that are able to derive regression coefficients and to evaluate a statistical value of the regression model assumed.

Whether or not any of the possible models are statistically significant is judged on the basis of several statistical significance parameters. Among them are: the correlation coefficient (R); the standard error of estimate (S_E); the value of the F-test of the overall significance (F); the values of t-test of significance of individual regressors (t); and the cross-correlation coefficients between the independent variables in the regression QSRR equation. Even if the values of these statistical parameters are within the acceptable range, one cannot exclude a chance correlation. This may result when too many variables are surveyed to correlate too few retention data.

Multivariate methods of data analysis, like discriminant analysis, factor analysis and principle component analysis, are often employed in chemometrics if multiple regression methods fail. The most popular chemometric method in QSRR is principle component analysis (PCA). By PCA one reduces the number of variables in a data set by finding linear combinations of those variables that explain most of the variability.

Commercially available software packages tabulate the component weights and the values of individual principal components. Plots of component weights for each variable (structural descriptor) are useful in QSRR analysis. Analogously, scatterplots for the first two principal components illustrate the distribution of objects (analytes) according to their inputs to the principal components.

There is an approach is QSRR in which principal components extracted from analysis of a large table of structural descriptors of analytes are regressed against the retention data in a multiple regression, i.e. principal component regression (PCR). The partial least squares (PLS) approach with cross-validation also finds application in QSRR.

DropBooks

Neural networks (NN) is a method of data analysis that emulates the brain's way of working. NNs are considered powerful tools and techniques for carrying out signal processing, modelling, forecasting and pattern recognition. A NN has its input neurons that load the system with descriptor values. Next, there are the hidden layers that weight and mix the incoming signals, and an output layer with neurons predicting the calibrated response values. The advantage of NNs lies in nonlinear transformations of signals occurring at every neuron. The NNs are trained to respond properly using a representative set of structural data and the corresponding retention parameters. The well-trained (but not an overtrained) NN predicts retention based on input information of an analyte.

Selection of Structural Descriptors

The translation of molecular structures into numerical descriptors is important not only in QSRR but also to many subdisciplines of chemistry and pharmacology.

A popular strategy for identifying structural parameters in QSRR analysis is to start from the accepted theories of chromatographic separation. Such structural parameters should quantify the abilities of analytes to take part in the postulated intermolecular interactions that determine chromatographic separations. Empirical or semiempirical structural parameters of analytes based on the solvatochromic comparison method and on linear solvation energy relationships (LSER) belong to that category of structural descriptors. Also, reliable predictions of retention have been demonstrated using the LFER-based experimental substituent or fragmental constants.

The structural descriptors that are commonly used in QSRR analysis are classified in **Table 1**.

Table 1 Structural descriptors in QSRR

Classification	*Descriptors*
Molecular bulkiness descriptors	Carbon number Molecular mass Refractivity Polarizability Van der Waals volume and area Solvent-accessible volume and area Total energy
Molecular geometry descriptors	Length-to-breadth ratio Moments of inertia Shadow area parameters
Physicochemical empirical and semiempirical parameters	Hammett constants Hansch constants Taft steric constants Hydrophobic fragmental parameters Solubility parameters Linear solvation energy relationship (LSER) parameters Partition coefficients Boiling temperatures pK_a values
Molecular polarity descriptors	Dipole moments Atomic and fragmental electron excess charges Orbital energies of HOMO and LUMO Partially charged areas Local dipoles Submolecular polarity parameters
Molecular topological descriptors	Molecular connectivity indices Kappa indices Information content indices Topological electronic index
Indicator variables	Zero-one indices
Ad hoc designed descriptors	

The structural descriptors related to molecular size may be related to the ability of an analyte to take part in nonspecific intermolecular interactions (dispersive interactions or London interactions) with the components of a chromatographic system. They are the factors most often found significant in QSRR analysis. The bulkiness parameters are decisive in the description of separations of closely congeneric analytes. For example, carbon number normally suffices to differentiate the members of a homologous series. On the other hand, when dealing with the set of analytes of the same size (e.g. isomers), they may appear not to be significant in QSRR analysis. This does not mean that dispersive interactions are meaningless for separation of congeners but just that they are closely similar, and hence the respective term in the QSRR equation apparently loses statistical significance.

What is more or less intuitively understood as molecular polarity of an analyte is difficult to quantify unequivocally. The descriptors of polarity are expected to account for differences among analytes regarding their dipole–dipole, dipole-induced dipole, hydrogen bonding and electron pair donor–electron pair acceptor (EPD–EPA) interactions. To find a good descriptor of these chemically specific interactions is difficult; the more so since changes in the polarity of an analyte also change its ability to take part in size-related interactions and affect analyte geometry.

Obviously, geometry-related or molecular shape parameters are difficult to quantify one-dimensionally. Single numbers reflecting molecular shape

differences are adequate only in the case of rigid and planar solutes. They become significant in QSRR equations if the range of analytes considered comprises compounds of similar size and polarity.

Physical meaning of the molecular graph-derived descriptors is never clear *a priori*. It is rather that good QSRR allow for assigning physical meaning to individual topological indices.

The empirical physicochemical parameters have good informative value for determining the mechanism of retention operating in a given chromatographic system. The problem is, however, the lack of such descriptors for the analytes of interest in actual QSRR studies.

Indicator values ('dummy variables') 0–1 are assigned depending on the presence or absence of a given structural feature in an analyte molecule. They serve to improve statistics but help occasionally to identify a structural descriptor of real physical significance.

The established structural descriptors listed in Table 1 seldom suffice to derive QSRR for the actual chromatographic data and often *ad hoc* descriptors have to be designed and included. QSRR analysis helps to test the predictive potency of the proposed structural descriptors, which may also appear suitable for deriving other kinds of structure–property relationships.

DropBooks

Prediction of Retention

Prediction of retention within homologous series is feasible owing to the linear relationships that are normally observed between retention parameters, log k, and carbon numbers of analytes, n (**Figure 2**). The slopes of lines, B, for various homologous series chromatographed under the same conditions are very similar, whereas the intercepts, A, may vary:

$$\log k = A + Bn \qquad [6]$$

Occasionally linear correlations are observed between retention parameters and molecular bulkiness descriptors of analytes that are not homologues. A good prediction of retention within a series of related nonpolar analytes, such as polyaromatic hydrocarbons (PAH) or alkylbenzenes, can be obtained using van der Waals volume as the structural descriptor.

The bulkiness descriptors can account for separation of analytes when dispersive interactions (London interactions) are the only interactions effective in a given chromatographic system, or when differences in polar interactions among analytes are not significant.

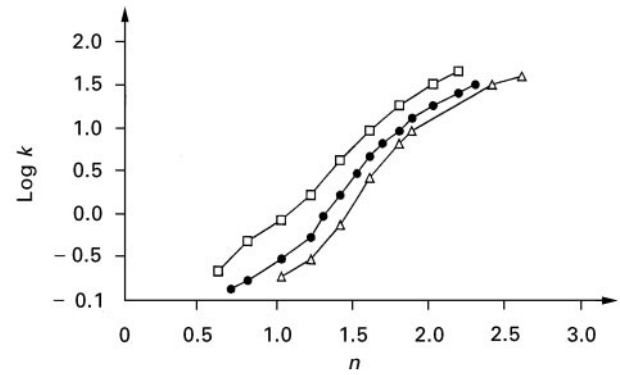


Figure 2 Plots of log k versus carbon number, n, of analyte for HPLC on a polyfunctional C_{18}-bonded silica with pure methanol eluent at 25°C: *n*-alkanes (□), methyl-*n*-alkanoates (●) and *n*-alkanols (○). (Reprinted from Tchapla A, Herson S, Lessellier E and Colin H (1993) General view of molecular interaction mechanisms in reversed-phase liquid chromatography. *Journal of Chromatography A* 656: 81–112, with permission of Elsevier Science.)

The ability of an analyte to take part in polar interactions is difficult to characterize by means of a single descriptor. Hence simple QSRR involving analyte polarity descriptors, e.g. dipole moments, are rare.

Normally in chromatography (excepting affinity chromatography) molecular shape effects on retention are of minor importance in comparison with the effects of molecular size and molecular polarity. In the case of planar/nonpolar PAH isomers, retention is linearly related to a shape descriptor (the degree of elongation of the analyte molecule).

There are numerous reports on good performance of the molecular connectivity index (Randić index) and its modifications in predicting retention of congeneric analytes, including isomers. The correlations are good when retention is on nonpolar stationary phases, but not when polar phases are involved. Whereas on the nonpolar phases the nonspecific dispersive interactions determine differences in retention among the analytes, the more specific polar interactions become discriminative in the case of polar phases (and polar analytes).

Using substituent electronic constants to derive simple QSRR with a real retention prediction ability has seldom succeeded. A wider application in that respect is found in Hansch substituent hydrophobic constants, π, and Rekker or Hansch–Leo fragmental hydrophobic constants, f. The sums of these constants (plus corrections due to intramolecular interactions) account well for retention in reversed-phase liquid chromatographic systems.

Regarding the latter systems, even better predictions are provided by an empirical parameter – the

logarithm of the *n*-octanol–water partition coefficient, log *P*. Another useful empirical retention predictor appears occasionally to be boiling point, T_b, for example the boiling point of isomeric hydrocarbons in the gas chromatography.

Prediction of retention of variously substituted derivatives of parent compounds in a given separation system can be based on the Martin rule:

$$\log k_S = \log k_P + \sum_{i=1}^{n} \tau_i \qquad [7]$$

In eqn [7] k_P is the retention parameter of a parent compound, k_S is the corresponding value for the derivative carrying n substituents and the summation represents the retention increments due to individual substituents τ_i. Having appropriate values for functional groups of interest, one needs only to determine retention of the parent structure to be able to calculate retention of a derivative. To get reliable predictions, correction factors are introduced in eqn [7] to account for mutual interactions between substituents (electronic, steric, hydrogen bonding). However, in the case of polyfunctional analytes, interactions between substituents make retention predictions of rather limited value.

A semiempirical description of reversed-phase HPLC systems, allowing for the prediction of the relative retention and selectivity within a series of analytes, has been developed by Jandera. The approach consists of determining the interaction indices and the structural lipophilic and polar indices. A suitable set of standard reference analytes is necessary to calibrate the retention (or selectivity) scale.

The multiparameter QSRR based on linear solvation energy relationships (LSER) possess a high predictive power regarding reversed-phase HPLC retention. The model developed by Abraham and co-workers to predict the *n*-octanol–water partition coefficient, log *P*, appears to be useful also in the case of log *k* from reversed-phase liquid chromatography:

$$\log k = c_0 + c_1 V_x + c_2 \pi_2^H + c_3 \sum \alpha_2^H + c_4 \sum \beta_2^H + c_5 R_2 \qquad [8]$$

In eqn [8] V_x is the so-called McGowan's characteristic volume, which can be calculated simply from molecular structure; π_2^H is the dipolarity/polarizability of the analyte, which can be determined through gas-chromatographic and other measurements; $\sum \alpha_2^H$ is the effective or summation hydrogen bond acidity; $\sum \beta_2^H$ is the effective or summation hydrogen bond basicity; and R_2 is an excess molar refraction, which can be obtained from refractive index measurements and is an additive quantity. The LSER-based structural descriptors are available for a large number of compounds.

Experimentally determined ionic radius, I_r, and energy of ionization, E_i, accompanied by atomic mass, A_m, produce a three-parameter regression equation predicting capillary electrophoretic mobility of metal cations. The QSRR equation indicates that atomic mass approximates to the retardation factors (negative input to mobility) whereas the ionic radius is an approximate measure of the effective charge on the analyte. Energy of ionization can play the role of a secondary, but significant, correction factor to the effective charge. Unfortunately, there are no good QSRR to predict the CE retention of organic analytes.

A typical multiparameter approach to predicting retention of an unknown compound based on structural features and chromatographic properties of other representative compounds consists of generating a multitude of analyte descriptors that are next regressed against retention data. The structural descriptors are usually derived by computational chemistry methods for the energy-optimized conformations. Software systems have been developed that produce and process hundreds of quantum chemical, molecular modelling, topological and semiempirical additive–constitutive descriptors after sketching the molecule on the computer. Observing all the rules and recommendations for meaningful statistics, the minimum number of descriptors (uncorrelated) is selected that are needed to produce a QSRR equation with a good predictive ability. The descriptors that eventually serve to predict retention of new analytes are sometimes of obscure physical meaning. For example, it is difficult to ascribe definite physical sense to such descriptors reported in predictive QSRR as 'the surface area of the positively charged portion of the molecule divided by the total surface area' or 'total entropy of the molecule at 300 K divided by the number of atoms'. Nevertheless, for several groups of compounds, prediction of retention by means of QSRR is reliable enough for identification purposes, especially when there is no better alternative. Exemplary predictive QSRR are for polychlorinated dibenzofurans and biphenyls, anabolic steroids, stimulants and narcotics used as doping agents, barbituric acid derivatives, polyaromatic and nitrated polyaromatic hydrocarbons, etc.

There are QSRR of useful predictive potency that comprise only physically interpretable terms. Reversed-phase HPLC retention of simple aromatic solutes on typical octadecyl silica columns has been related to a molecular bulkiness descriptor (total energy), a polarity descriptor (local dipole) and the energy of the

highest occupied molecular orbital of analytes. Good prediction of liquid chromatographic retention of about 50 aromatic acids was realized using as regressors the calculated theoretical logarithm of the *n*-octanol–water partition coefficient ($\log P$), the dipole moment, the principal ellipsoid axes, the sum of the charges on the oxygen and nitrogen atoms, the energy of the highest occupied molecular orbital (HOMO) and the electrophilic superdelocalizability for the aromatic carbon atom.

In **Figure 3** is illustrated the predictive performance of QSRR for 216 HPLC retention data points. The points are for 36 analytes chromatographed in six eluents on a diol stationary phase. The eluents were heptane containing 0.5% of tetrahydrofuran, dioxane, ethanol, propanol, octanol and dimethylformamide. In Figure 3 the $\log k$ data experimentally measured are plotted against the values predicted by eqn [9]:

$$\begin{aligned}\log k = {} & 0.100\ \text{polarizability (analyte)} \\ & - 0.400 \log P\ \text{(analyte)} \\ & - 0.330 E_{\text{HOMO}}\ \text{(analyte)} \\ & + 1.106 E_{\text{HOMO}}\ \text{(eluent)} \\ & + 0.401 E_{\text{LUMO}}\ \text{(eluent)}\end{aligned} \quad [9]$$

with the values $n = 216$, $R = 0.97$, $s = 0.097$ and $F = 655$. In this equation n is the number of data points used to derive regression equation, R is the multiple correlation coefficient, s is the standard error of estimate and F is the value of F-test of significance; E_{LUMO} denotes energy of the lowest unoccupied molecular orbital and E_{HOMO} is energy of the highest occupied molecular orbital.

Figure 3 reflects realistically the actual predictive power of QSRR. The predictive QSRR equations normally hold within the family of analytes for which they were derived and may be used for tentative identification of chromatographic peaks.

In recent years a three-dimensional quantitative structure–biological activity relationship method known as comparative molecular field analysis (CoMFA) has been applied to construct a 3D-QSRR model for prediction of retention data. The CoMFA 3D-QSRR model is obtained by systematically sampling the steric and electrostatic fields surrounding a set of analyte molecules and then correlates the differences in these fields to the corresponding differences in retention.

Several reports have recently appeared on predictions of retention data from structural descriptors by means of neural networks (NN). By now the predictions provided by NN are of similar reliability to those obtained from regression models.

QSRR and Molecular Mechanisms of Retention

The QSRR equations that comprise physically interpretable structural descriptors can be discussed in terms of the molecular mechanisms involved in the chromatographic process. There is evidence that different structural parameters of analytes account for retention differences in GC on polar versus nonpolar stationary phases. Also, the structural descriptors in

DropBooks

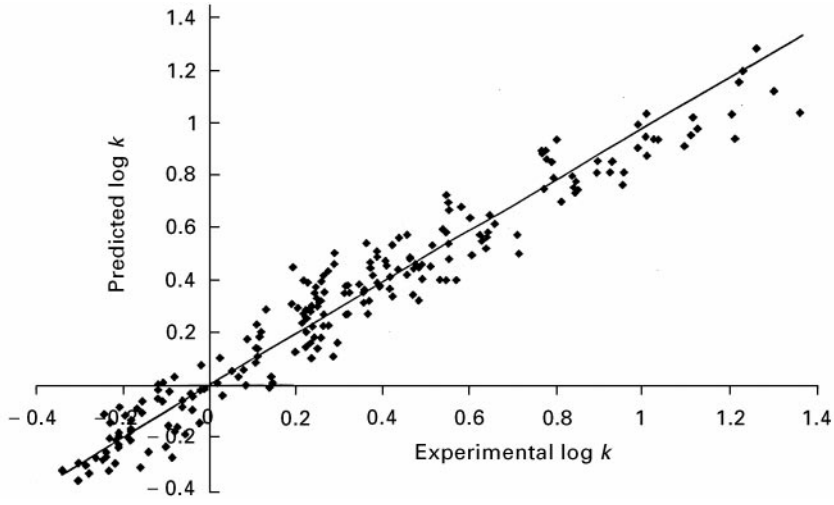


Figure 3 Plot of log *k* predicted by eqn [9] against experimental data determined on a diol column for 36 chalcone derivatives with heptane eluent containing 0.5% tetrahydrofuran, dioxane, ethanol, propanol, octanol or dimethylformamide. (Reprinted with permission from Azzaoui K and Morin-Allory L (1996) Comparison and quantification of chromatographic retention mechanisms on three stationary phases using structure–retention relationships. *Chromatographia* 42: 389–395. Copyright Friedrich Vieweg & Sohn.)

QSRR equations that are valid for normal and for reversed-phase HPLC systems are different. In the case of apparently similar chromatographic systems, the differences in retentive properties of stationary phases may be reflected by the magnitude of the regression coefficients for analogous descriptors. Comparative QSRR studies are especially valuable when new chromatographic phases are introduced.

A general rule is that QSRR equations are characterized by two kinds of structural descriptors: one that accounts for the bulkiness or size of an analyte and one that encodes its polar properties. Size descriptors are always significant in GC on nonpolar phases and in reversed-phase HPLC, whereas the significance of polar descriptors increases as polarity of both the stationary phases and the analytes increases.

Publications give evidence that in GC on polar phases and in normal-phase (adsorption) liquid chromatography (HPLC and TLC) the chemically specific, molecular size-independent intermolecular interactions are assumed to play the main retention-determining role. For example, the HPLC retention parameters determined for substituted benzenes on porous graphitic carbon are described by QSRR equations comprising polarity descriptors but no bulk descriptors. Because, in general, it is difficult to quantify polarity properties precisely, the QSRR for GC on polar phases and for normal-phase HPLC are usually of lower quality then for GC on nonpolar phases and in reversed-phase liquid chromatography.

QSRR differentiate in a quantitative (statistical) manner stationary phase materials of different chemical nature. However, when the stationary phases that belong to the same chemical class are compared, such as hydrocarbon-bound silicas for reversed-phase HPLC, the results obtained are ambiguous.

The proper QSRR strategy aimed at objective characterization of differences in retentive potency of individual chromatographic systems should employ a well-designed set of test analytes. The analytes should be selected in such a way that, within the test set, the intercorrelations are minimized among the individual analyte structural descriptors. At the same time, the selection of test analytes should provide a wide range and even distribution of individual structural descriptor values. In addition, the series of analytes should be large enough to assure statistical significance of the QSRR equations but not too large so as to remain experimentally manageable.

Often the retention parameters of test analytes are first linearly regressed against the reference $\log P$ values from the *n*-octanol–water partition system. Good correlations obtained are usually interpreted as evidence of a partition mechanism operating in the chromatographic system under study.

Several QSRR studies aimed at comparison of retention mechanisms on individual alkyl silica reversed-phase materials for HPLC have employed LSER-based analyte parameters. It was observed generally that the most important analyte parameters that influence retention are bulkiness-related (molar volume, molar refraction) and hydrogen bonding basicity, but not hydrogen bonding acidity. The analyte dipolarity/polarizability appeared to be a minor but often significant factor. However, on poly(st yrene–divinylbenzene) stationary phases the dipolarity/polarizability term provides an important positive input to QSRR.

The results of QSRR studies in which eqn [8] was applied to the retention parameters, $\log k_w$, from measurement on alkyl silica phases with methanol–water and acetonitrile–water eluents are instructive. The most significant parameters appeared to be hydrogen bond basicity (β_2^H) and McGowan volume (V_X) of analytes. The third significant parameter in QSRR equations is either dipolarity/polarizability (π_2^H) in the case of methanolic eluents or hydrogen bond acidity (α_2^H) in the case of acetonitrile-modified mobile phases.

The rationalization of these results might be as follows. The dispersive interactions of analytes (characterized by V_X) and hydrogen bonding interactions in which an analyte molecule is a hydrogen-bond acceptor (characterized by $\sum\beta_2^H$) significantly affect the retention of analytes in both water–methanol/stationary phase and water–acetonitrile/stationary phase equilibrium systems. However, in methanolic systems the third significant factor determining equilibrium is the ability of an analyte molecule to be preferentially attracted by polar molecules of methanol owing to the dipole–dipole and dipole-induced dipole interactions (characterized by π_2^H). In the systems containing acetonitrile, the π_2^H descriptor becomes insignificant in QSRR equations. Instead, the ability of the analyte to be preferentially attracted by the eluent owing to hydrogen bonding in which the analyte serves as a hydrogen bond donor (characterized by $\sum\alpha_2^H$) becomes more significant. The well-known hydrogen bond acceptor properties of acetonitrile manifest themselves in eqn [8] as a retention-decreasing term $k_4\sum\alpha_2^H$ with a negative value of the k_4 regression coefficient.

Most readily interpretable would appear to be the molecular mechanism of retention in terms of QSRR equations comprising the parameters of analytes obtained from molecular modelling. One can easily assign physical meaning to van der Waals surface area or solvent-accessible molecular surface area (*SAS*) as differentiating the strength of dispersive interactions between the analyte and the molecules forming

chromatographic systems. Dipole moment (μ) should also account for differences among analytes regarding their dipole–dipole or dipole-induced dipole interactions. Energies of the lowest unoccupied molecular orbital (E_{LUMO}) and the highest occupied molecular orbital (E_{HOMO}) should explain the differences in the tendency of analytes to take part in the charge transfer interactions. Yet reliable QSRR employing these structural descriptors are rare and hold only for selected sets of analytes.

In QSRR concerning reversed-phase HPLC retention parameters, the net positive effects on retention are due to the analyte bulkiness descriptors. The dispersive attractions of an analyte are stronger with the bulky hydrocarbon ligand of the stationary phase than with the small molecules of aqueous eluent. The net effect on retention provided by dipole moment (or its square) is negative. This is because the dipole–dipole and dipole-induced dipole attractions are stronger between the polar (polarized) analyte and polar molecules of eluent than between the same analyte and the nonpolar hydrocarbon ligand of the stationary phase. Unfortunately, these types of QSRR are not precise enough to differentiate individual alkyl silica stationary phase materials in a quantitative, statistically significant manner. They are significant enough, however, to reflect the differences in retention mechanism operating in the reversed-phase and in the normal-phase HPLC systems or in GC on nonpolar and polar phases.

Factorial methods of data analysis (principal component analysis, correspondence factor analysis, spectral mapping analysis) provide classification of stationary phases based on retention data determined for short series of test analytes. Among the commercially available materials for HPLC those can be selected that possess closely similar retention properties. Also, a stationary phase with clearly distinctive properties can be identified, which can be useful for specific method development.

Chromatographic Methods of Determination of Hydrophobicity

Hydrophobicity or lipophilicity is understood to be a measure of the relative tendency of an analyte 'to prefer' a nonaqueous over an aqueous environment. The partition coefficients of the substances may differ if determined in different organic–water solvent systems but their logarithms are often linearly related. Octanol–water is a reference system that provides the most commonly recognized hydrophobicity measure: the logarithm of the partition coefficient, log P. The standard 'shake-flask' method for determining partition coefficients in liquid–liquid systems has several disadvantages. Having appropriate QSRR, the chromatographic data can be used to predict log P. Many good correlations of reversed-phase liquid chromatographic (HPLC or TLC) parameters with log P have been reported for individual chemical families of analytes and chromatographic methods for assessing the hydrophobicity of drugs and environmentally important substances have officially been acknowledged and included in the *OECD Guidelines for Testing Chemicals*.

On the other hand, the partition chromatographic systems are not identical with the *n*-octanol–water partition system. Each chromatographic system produces an individual scale of hydrophobicity. Hence attempts to reproduce log P by means of liquid chromatography are only partially successful. Centrifugal countercurrent chromatography (CCCC) provides a better chance of mimicking log P but the inconvenience of this method and the need for special equipment hinder its wider application.

The versatility of chromatographic methods of hydrophobicity assessment can be attributed to the use of organic modifiers in aqueous eluents. Normally, the retention parameters determined at various organic modifier–water (buffer) compositions are extrapolated to zero organic modifier content. The extrapolated parameters ($\log k_w$ from HPLC and R_M^0 from TLC) depend on the organic modifier used.

Alkyl silica stationary phases and methanol–water eluent are most often used in hydrophobicity studies. The problem with these phases is that the hydrophobicity of nonionized forms of organic bases cannot be determined because of the chemical instability of silica-based materials at higher pHs (above about pH 8). Also, specific interactions of analytes with the free silanols of alkyl silicas disturb partition processes.

The limitations of standard reversed-phase materials have been partially overcome by introducing modern specially deactivated hydrocarbon-bonded phases, immobilized on alumina or zirconia supports and on polymeric materials. Using the latter two types of stationary phase materials one can determine HPLC retention factors under acidic, neutral and alkaline conditions. That way a universal, continuous chromatographic hydrophobicity scale can be constructed, as is the standard log P scale.

Hydrophobic properties of xenobiotics are assumed to affect their passive diffusion though biological membranes and binding to pharmacological receptors. If the hydrophobicity measuring system is to model a given biological phenomenon, then similarity of the component entities is a prerequisite.

DropBooks

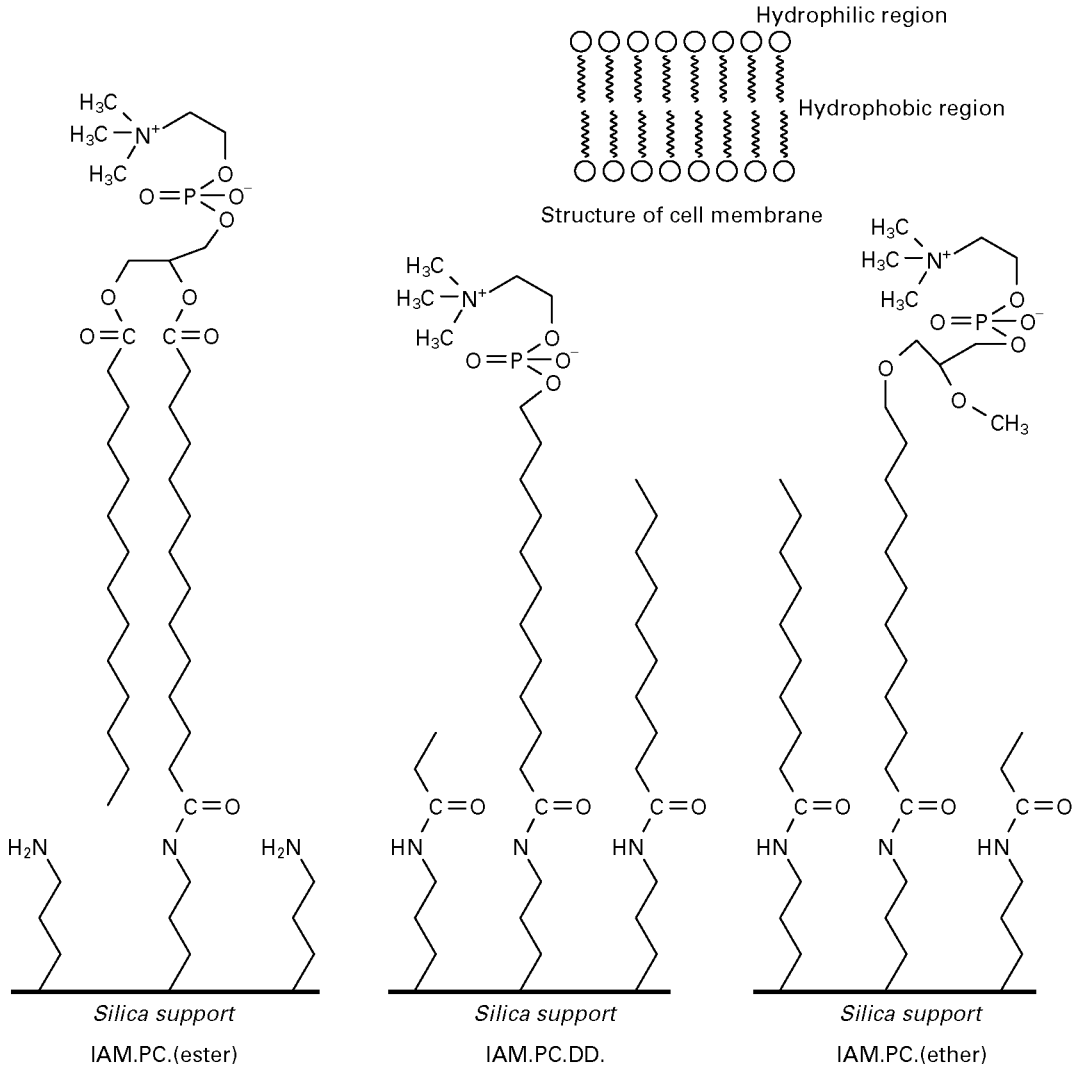


Figure 4 Chemical structures of ligands of three types of immobilized artificial membrane (IAM) columns of Pidgeon (Liu H, Ong S, Glunz L and Pidgeon C (1995) Predicting drug–membrane interactions by HPLC: structural requirements of chromatographic surfaces. *Analytical Chemistry* 67: 3550–3557. Copyright 1995 American Chemical Society.) and a schematic model of a biological membrane.

Hence the partition system to model the transport through biological membranes should be composed of an aqueous phase and an organized phospholipid layer. The immobilized artificial membranes (IAM) introduced by Pidgeon as a packing material for HPLC (**Figure 4**) appear to be reliable and convenient models of natural membranes.

Correlations between $\log k$ data determined on IAM-type columns and $\log P$ values are generally not high nor are the correlations between $\log k$ from IAM columns and $\log k_w$ determined by liquid chromatography employing standard stationary phase materials. This means that retention data determined on IAM columns contain information on the properties of analytes that is distinct from that provided by the *n*-octanol–water system and by the hydrocarbon–silica reversed-phase columns. There is evidence that the hydrophobicity characteristics provided by IAM columns are better suited for modelling the pharmacokinetics of drug processes.

Retention Parameters in Predicting Bioactivity of Analytes

Biological processes of drug absorption, distribution, excretion and drug–receptor interaction are dynamic in nature as are the analyte's distribution processes in chromatography. The same fundamental intermolecular interactions determine the behaviour of chemical compounds in both biological and chromatographic environments. Modern techniques and procedures of HPLC and CE allow for inclusion

of biomolecules as active components of separation systems and QSRR processing of appropriate sets of chromatographic data can reveal systematic information regarding the xenobiotics studied. This information can be used to elucidate molecular mechanisms of pharmacological action and to facilitate rational drug design.

It is often sufficient to identify and employ a specific chromatographic system yielding hydrophobicity values of analytes best conforming to log P data. Normally, chromatographic systems that produce retention parameters less correlated to log P are discarded but information extracted from diversified retention data may be more appropriate for prediction of pharmacological properties of analytes than information based on an individual hydrophobicity scale. To extract meaningful information from diversified (yet often highly mutually intercorrelated) sets of data, multivariate chemometric methods of data analysis are employed. Large matrices of retention data determined for test series of analytes in many chromatographic systems differing in type of stationary and/or mobile phases, are processed by factorial methods, usually by principal component analysis (PCA). If two to three extracted abstract factors (principal components) account for most of the variability in a large set of retention data then the distribution of test analytes can be presented graphically. Clustering of analytes owing to similarity of their chromatographic behaviour in diverse separation systems is usually observed. If that clustering agrees with the pharmacological classification of the test agents, then recalculations are done after including the retention data for drug candidates. Indications on potential pharmacological activity of new analytes can be obtained even before biological experiments. This approach can facilitate preselection of drug candidates, especially among a multitude of compounds produced by combinatorial chemistry. The challenge is to design and select the chromatographic systems yielding retention data of sufficient classification potential.

Figure 5 shows the distribution of drugs belonging to several pharmacological classes on the plane determined by the first two principal components, which together account for 81.5% of the variance in the retention data measured in eight HPLC systems. The HPLC systems employed different stationary phases (standard and specially deactivated hydrocarbon bonded silicas, polybutadiene-coated alumina, immobilized artificial membranes and immobilized α_1-acid glycoprotein). Methanol–buffer eluents of varying composition and pH were used. The clustering of analytes is consistent with their established pharmacological classification. Also, the partial overlap of individual clusters is interpretable in terms of partially overlapping pharmacological properties of individual drugs.

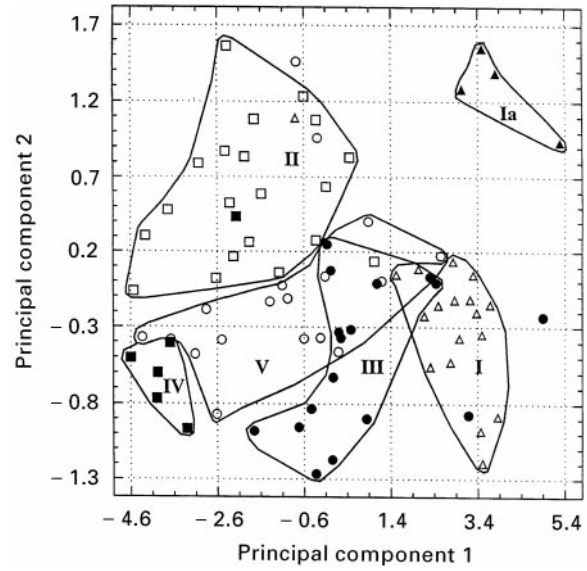


Figure 5 Pharmacologically consistent distribution scatterplot of drug classes on the plane determined by two first principal components extracted from a 8×83 (drugs $\times$ HPLC systems) matrix of diversified retention data. Roman numbers denote: I, psychotropic drugs; Ia, inactive phenothiazines; II, β-adrenolytics; III, histamine H_1 receptor antagonists; IV, histamine H_2 receptor antagonists; V, drugs binding to α-adrenoceptors. (Reprinted from Nasal A, Buciński A, Bober L and Kaliszan R (1997) Prediction of pharmacological classification by means of chromatographic parameters processed by principal component analysis. *International Journal of Pharmaceutics* 159: 43–55, with permission of Elsevier Science.)

There are individual processes of drug action that are satisfactorily modelled by HPLC on immobilized artificial membrane (IAM) columns. QSRR equations have been reported predicting several pharmacokinetic parameter of β-adrenolytic drugs from their log k parameters determined on IAM columns. Good predictions by means of log k_{IAM} have also been reported regarding antihaemolytic activity of phenothiazine neuroleptics. The human skin permeation of steroids also correlates better with log k_{IAM} than with log P.

The log k_{IAM} alone will not suffice to predict binding of basic drugs to the serum protein, α_1-acid glycoprotein (AGP). However, combining that parameter with atomic excess charge on aliphatic nitrogen, N_{ch}, and a size parameters, S_T, in a multiple regression equation results in a good prediction of AGP binding. The S_T parameter is the area of a triangle having one vertex on the aliphatic nitrogen and the two remaining vertices on the extremely positioned atoms in the drug molecule (**Figure 6**). The QSRR equation has

DropBooks

the form:

$$\log k_{AGP} = 0.6577\,(\pm 0.0402)\log k_{IAM} + 3.342(\pm 0.841)N_{ch} - 0.0081(\pm 0.0030)S_T + 1.688(\pm 0.245) \quad [10]$$

with the values $n = 49$, $R = 0.929$, $s = 0.163$, $F = 92$ and $p < 10^{-5}$.

Equation [10] may be useful as a first approximation to the relative binding of a drug to AGP without the need to perform biochemical experiments. It can help to identify structural features of the binding site of basic drugs on AGP (**Figure 6**). The site can be modelled as a conical pocket. Its internal surface contains hydrophobic regions at the base of the cone and an anionic region close to the apex of the cone. Protonated aliphatic nitrogen guides drug molecules towards the anionic region. Hydrophobic hydrocarbon fragments of the interacting drugs provide anchoring in the hydrophobic regions of the binding site. There is a steric restriction for the molecule to plunge into the binding site.

QSRR analysis of HPLC data determined on an immobilized human serum albumin (HSA) column helps to suggest the topography of two binding sites of different affinity to benzodiazepine enantiomers. Also, the mechanism of interaction of phenothiazine neuroleptics with melanin can be rationalized by means of QSRR analysis of HPLC retention data. Another QSRR study concerns interactions of drugs with immobilized keratin and collagen.

In general, QSRR analysis of retention parameters determined on immobilized biomacromolecules can yield reliable predictions of activity and identification of the required binding structural properties of drugs and drug candidates. The approach appears especially promising now that biotechnologically produced pharmacological receptors are becoming available.

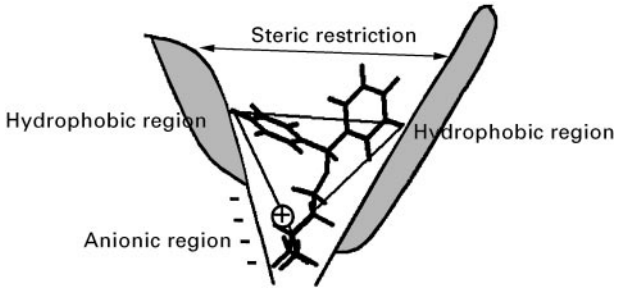


Figure 6 Mode of binding of the organic base drugs derived from QSRR analysis of HPLC data determined on an immobilized α_1-acid glycoprotein column. (Adapted with permission from Kaliszan R, Nasal A and Turowski M (1995) Binding site for basic drugs on α_1-acid glycoprotein as revealed by chemometric analysis of biochromatographic data. *Biomedical Chromatography* 9: 211–215. Copyright John Wiley & Sons Limited.)

Concluding Remarks

In 1991 Giddings wrote 'Because pure theory is impractical, progress in understanding and describing molecular equilibrium between phases requires a combination of careful experimental measurements and correlations by means of empirical equations and approximate theories'. This has been realized in a systematic manner over a period of 20 years through QSRR analysis. During that time a consistent research startegy has been developed and established within the area. Easy access to computers equipped with advanced statistics and molecular modelling software has ensured rapid progress and engendered a wide interest in QSRR among chromatographers and other specialists.

QSRR are employed by analytical chemists to help identify unknown members of individual classes of analytes of pharmacological, toxicological, environmental or chemical interest. At the same time, QSRR of good retention prediction capability help to identify structural descriptors for analytes that provide acceptable estimates of properties other than chromatographic ones. In this way, chromatographic systems allowing for fast and convenient evaluation of analyte hydrophobicity can be identified. Also, QSRR confirm the suitability of the LSER-based descriptors for property predictions.

Well-designed QSRR studies are helpful in identifying the structural features within a family of analytes that affect retention in a given separation system. In this way molecular mechanisms of retention may be explained. With a designed test series of analytes the QSRR derived for retention data determined in individual separation systems provide objective, numerical characteristics for these separation systems. This is especially useful for quantitatively comparing retention properties of various stationary phase materials.

Chromatographic retention data can be employed to predict pharmacological properties of analytes. By employing chromatographic systems comprising biomacromolecules, large amounts of data can be obtained that reflect differences among analytes with regard to their interactions with given biomacromolecules. These data can be used to derive QSRR explaining the mechanism of drug–biomacromolecule interactions. In effect, the topography of binding sites for drugs on individual biomacromolecules can be characterized. By employing

biotechnologicallly acquired pharmacological receptor proteins to generate drug–receptor interaction data and by applying QSRR analysis, the preselection of drug candidates can be facilitated and experiments on animals reduced.

See also: **II/Chromatography: Liquid:** Mechanisms: Reversed Phases.

Further Reading

Carr PW, Martire DE and Snyder LR (eds) (1993) The retention process in reversed-phase liquid chromatography. *Special Volume of Journal of Chromatography A* 656: 1–618.

Forgács E and Cserháti T (1997) *Molecular Bases of Chromatographic Separations*. Boca Raton, FL: CRC Press.

Giddings JC (1991) *Unified Separation Science*. New York: Wiley.

Jinno K (ed.) (1997) *Chromatographic Separations Based on Molecular Recognition*. New York: Wiley-VCH.

Jurs PC (1996) *Computer Software Applications in Chemistry*, 2nd edn. New York: Wiley.

Kaliszan R (1987) *Quantitative Structure–Chromatographic Retention Relationships*. New York: Wiley.

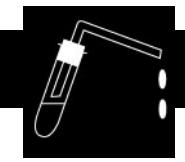
Kaliszan R (1997) *Structure and Retention in Chromatography. A Chemometric Approach*. Amsterdam: Harwood Academic Publishers.

Kier LB and Hall LH (1986) *Molecular Connectivity in Structure–Activity Analysis*. Letchworth: Research Study Press.

Pliška V, Testa B and van de Waterbend H (eds) (1996) *Lipophilicity in Drug Action and Toxicology*. Weinheim: VCH.

Smith RM (1995) *Retention and Selectivity in Liquid Chromatography*. Amsterdam: Elsevier.

REACTIVE DISTILLATION



S. M. Mahajani, Monash University, Clayton, Victoria, Australia
S. P. Chopade, Michigan State University, East Lansing, MI, USA

Copyright © 2000 Academic Press

DropBooks

Introduction

Reactive distillation is a combination of separation and reaction in a single process. Commercial reactive distillation processes for the manufacture of methyl *t*-butyl ether (MTBE) and methyl acetate were successfully commissioned in 1981 and 1983, respectively. These processes have a distinct edge over their conventional predecessors. The reactive distillation process is particularly advantageous in the case of reversible reactions where the conversion is limited by thermodynamic equilibrium. Some of the important benefits of reactive distillation are: reduced capital cost; employment of low mole ratios of reactants; energy saving owing to utilization of the heat of reaction; and automatic temperature control and elimination of hot spots. The commercial process of MTBE manufacture has shown that heterogeneous catalysts such as ion exchange resins can be advantageously used in reactive distillation columns. Innovative techniques of confining the small size resin particles (0.3–2 mm) in the column, allowing efficient solid–liquid contact and high void fraction, have been developed by CDTech, Sulzer, Koch Engineering and BASF. An alternative approach is to prepare a catalyst in the form of conventional column packing and pack it directly into the reactive distillation column.

Recognizing the potential of reactive distillation for a particular process is a difficult task, as not all the reactions can be conducted effectively in this way. Once its potential has been identified, the next step is to design the reactive distillation column for the required task. The simultaneous existence of multiple processes such as mixing, mass transfer and reaction are involved, and the design method requires thorough knowledge of both chemical and physical equilibria as well as the reaction kinetics. Graphical representations of liquid phase compositions, called residue curve maps or distillation maps, are commonly used to analyse the reactive distillation process. Though some efforts have been made to study the underlying theory of the design method, the work is still at its preliminary stage. Another approach to understanding the behaviour of this process is to perform computer simulations and predict the performance of a column of known configuration.

In this article the important aspects of commercial reactive distillation processes of MTBE and methyl acetate manufacture are described in detail. Recent trends in the experimental and theoretical investigations in this area are also outlined. The potential importance of reactive distillation in some industrial